The Simplified Component Technique: An Alternative to Rotated Principal Components

Author(s): Ian T. Jolliffe and Mudassir Uddin

Source: *Journal of Computational and Graphical Statistics*, Vol. 9, No. 4 (Dec., 2000), pp. 689-710
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of America
Stable URL: https://www.jstor.org/stable/1391088

The Simplified Component Technique: An Alternative to Rotated Principal Components

Ian T. JOLLIFFE and Mudassir UDDIN

It is fairly common, following a principal component analysis, to rotate components in order to simplify their structure. Here, we propose an alternative to this two-stage procedure which involves only one stage and combines the objectives of variance maximization and simplification. It is shown, using examples, that the new technique can provide alternative ways of interpreting a dataset. Some properties of the technique are investigated using a simulation study.

Key Words: Interpretation; Rotation; Simple structure.

1. INTRODUCTION

Principal component analysis (PCA) is primarily a dimension-reduction technique, which takes observations on p correlated variables and replaces them by uncorrelated variables which successively account for as much as possible of the variation in the original variables. These uncorrelated variables are the principal components (PCs), and they are linear combinations of the original variables. Typically using only the first m PCs $(m \ll p)$ instead of the p variables will involve only a small loss of variation.

One problem with using PCA, rather than the alternative strategy of replacing the p variables by a subset of m of the original variables, is interpretation. Each PC is a linear combination of all p variables, and we wish to decide which variables are important, and which are unimportant, in defining each of our m components. This can be done in a number of ways. Cadima and Jolliffe (1995) examined a number of possibilities including regression of each PC on the variables. Hausman (1982) and Vines (2000) considered modifications of PCA in which the values that the loadings can take are restricted, in Hausman's case to $-1, 0$, and 1. Jolliffe and Uddin (in press) introduced a constrained version of PCA in which some loadings are driven to zero. Here we concentrate modifying the traditional approach, based on rotation, to simplification of loadings.

It is fairly common practice in some fields of application—for example, atmospheric science (Richman 1986, 1987)—to rotate the loadings found in a PCA. This idea comes

Ian T. Jolliffe is Professor, Department of Mathematical Sciences, University of Aberdeen, Meston Building, King's College, Aberdeen AB24 3UE, Scotland, UK (E-mail: itj@maths.abdn.ac.uk). Mudassir Uddin is Lecturer, Department of Statistics, University of Karachi, Karachi-75270, Pakistan (E-mail: mudassir2000@hotmail.com).

Journal of Computational and Graphical Statistics, Volume 9, Number 4, Pages 689–710

Table 1. Definitions of Variables in Rencher's (1995) Meteorological Dataset

Variable	*Definition*
x_1	maximum daily air temperature
x_2	minimum daily air temperature
x_3	integrated area under daily air temperature curve
x_4	maximum daily soil temperature
x_5	minimum daily soil temperature
x_6	integrated area under soil temperature curve
x_7	maximum daily relative humidity
x_8	minimum daily relative humidity
x_9	integrated area under daily humidity curve
x_{10}	total wind, measured in miles per day
x_{11}	evaporation

from factor analysis, and proceeds by post-multiplying the $(p \times m)$ matrix of PC loadings by a (usually orthogonal) matrix to make the loadings "simple." Simplicity is defined by one of a number of criteria (e.g., varimax, quartimax) which quantify the idea that loadings should be near zero or near ± 1, with as few as possible intermediate values. Rotation takes place within the subspace defined by the first m PCs. Hence all the variation in this subspace is preserved by the rotation, but it is redistributed among the rotated components, which no longer have the successive maximization property. We illustrate the procedure with a simple example, given by Rencher (1995), which consists of 46 daily measurement on 11 meteorological variables. Table 1 lists the variables, and Table 2 and Figure 1 give loadings for PCs, for rotated PCs, for $m = 4$, using the well-known varimax rotation criterion, and for simplified components defined later. The unrotated analysis has nontrivial loadings on 9 of the 11 variables in its first PC, which are reduced to 6 after rotation. Similar simplifications occur in later components.

Rotation of principal components has a number of drawbacks (Jolliffe 1987, 1989, 1995). We defer detailed discussion of these until later in the article, when we compare the properties of rotated PCs with the components produced by a new technique. This simplified component technique (SCoT) has one stage, rather than the two needed in rotated PCA, and each simplified component optimizes a criterion combining the desirable properties of large variance and simplicity. Successive components are constrained to be orthogonal to, or uncorrelated with, one another. SCoT is defined in Section 2 and illustrated with examples in Section 3. Comparisons are made with rotated PCA in both Sections 2 and 3. Computational aspects of SCoT are discussed in Section 4. A simulation study in which some properties of SCoT are described in Section 5, and further discussion and concluding remarks are given in Section 6.

2. THE SIMPLIFIED COMPONENT TECHNIQUE

2.1 PRINCIPAL COMPONENT ANALYSIS

To establish notation we formally define PCA. Let $\mathbf{x}$ be a vector of p random variables with covariance matrix Σ. The kth principal component, for $k = 1, 2, \ldots, p$, is the linear function of $\mathbf{x}$, $\boldsymbol{\alpha}_k'\mathbf{x}$, which maximizes $\mathrm{var}[\boldsymbol{\alpha}_k'\mathbf{x}] = \boldsymbol{\alpha}_k'\Sigma\boldsymbol{\alpha}_k$ subject to $\boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = 1$, and, (for $k > 1$), $\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_k = 0, j < k$.

Often PCA is done for variables standardized to each have unit variance, so that Σ becomes the correlation matrix for the original variables. In practice, we work with a $(n \times p)$ data matrix $\mathbf{X}$, and the *sample* covariance/correlation matrix, $\mathbf{S}$, for these data. We denote the value of the kth PC for the ith observation (row of $\mathbf{X}$) by $\mathbf{a}_k'\mathbf{x}_i$ in this case.

2.2 ROTATION

Suppose that we have decided to retain and rotate m PCs and that $\mathbf{A}$ is the $(p \times m)$ matrix whose kth column is $\mathbf{a}_k, k = 1, 2, \ldots, m$. Rotating the first m components is achieved by post-multiplying $\mathbf{A}$ by a matrix $\mathbf{T}$, to obtain rotated loadings $\mathbf{B} = \mathbf{AT}$. The choice of $\mathbf{T}$ is determined by whichever rotation criterion we use. For example, for the commonly used varimax rotation criterion, $\mathbf{T}$ is an orthogonal matrix chosen to maximize

$$S(\mathbf{B}) = \frac{1}{p^2} \sum_{k=1}^{m} \left[m \sum_{j=1}^{p} b_{jk}^4 - \left(\sum_{j=1}^{p} b_{jk}^2 \right)^2 \right], \tag{2.1}$$

where b_{jk} is the (j,k)th element of $\mathbf{B}$ (Krzanowski and Marriott 1995).

The idea behind the varimax criterion is to simplify the structure of the loadings by maximizing the variance of squared loadings within each column of $\mathbf{B}$. This drives the loadings towards zero or ± 1; alternative rotation criteria attempt to achieve similar objectives, but using a variety of other specific definitions. Concentrating the loadings close to zero or ± 1 is not the only possible definition of simplicity. For example, if all loadings were equal, that is also simple but in a completely different way. Building a criterion to take into account both types of simplicity would be difficult, and hence we stick to the usual concept, as typified by varimax.

2.3 THE SIMPLIFIED COMPONENT TECHNIQUE

The simplified component technique (SCoT) reduces the two stages of rotated PCA into one step in which we successively attempt to find linear combinations of the p variables that maximize a criterion which consists of the variance, supplemented by a penalty function which pushes the linear combination towards simplicity. Let $\mathbf{c}_k'\mathbf{x}_i$ be the value of the kth SC for the ith observation and suppose, for example, that we define simplicity in terms of the varimax criterion $S(\mathbf{c}_k)$, as defined in Equation (2.1), but for the special case where $m = 1$.
If

$$V(\mathbf{c}_k) = \mathrm{var}(\mathbf{c}_k'\mathbf{x}), \tag{2.2}$$

then the SCoT successively maximizes

$$V(\mathbf{c}_k) + \phi S(\mathbf{c}_k), \tag{2.3}$$

subject to $\mathbf{c}_k'\mathbf{c}_k = 1$ and, (for $k \geq 2$) either $\mathbf{c}_j'\mathbf{c}_k = 0$, or $\mathbf{c}_j'\mathbf{S}\mathbf{c}_k = 0, j < k$. Here ϕ is a simplicity/complexity parameter.

Remark 1. There are links between SCoT and other statistical techniques. Penalty functions have been used in several applications in statistics. In many cases, the penalty function is a roughness penalty, included to increase smoothness (see, e.g., Titterington 1985; Rice and Silverman 1991; Green and Silverman 1994) rather than a complexity penalty which targets simplicity. If $\phi = 0$, we simply obtain PCA, while as $\phi \to \infty$, we will eventually get each SC containing only one of the original variables.

There are also connections with projection pursuit. The latter finds "interesting" low (m-) dimensional projections of a high (p-) dimensional dataset. "Interesting" can mean a variety of things (Huber 1985; Jones and Sibson 1987), but often indicates structures such as clusters or outliers. If "high variance" can be thought of as "interesting," PCA is a special case of projection pursuit. Indeed if the main source of variation in a dataset is that between a set of clusters, PCA will find such clusters. Huber (1985) suggested that some types of factor analysis are also special cases of projection pursuit. However, the rotation associated with factor analysis is concerned with simplifying the interpretation of the axes defining the m-dimensional projections, rather than seeking projections with interesting structures. Morton (1989) modified projection pursuit by incorporating a penalty function based on a simplicity index, and it is this approach, in the context of PCA, that we have adopted.

Remark 2. The one-step procedure of SCoT cannot give equivalent results to two-stage rotated PCA. The rotated PCs remain within the subspace obtained by the m retained PCs; that is, the m-dimensional subspace with *maximum* variance. The introduction of the penalty function necessarily takes us outside this subspace, so SCoT sacrifices some variance in its search for simplicity. There is no obvious sense in which the m-dimensional space defined by SCs is globally optimal.

Remark 3. The one-step procedure operates successively, so a decision to increase the number of retained components from m to $m + 1$ leaves the first m components unchanged. With rotated PCs a change from m to $m + 1$ may change the nature of *all* the rotated PCs.

Remark 4. Computationally, finding PCs followed by rotation is straightforward. This is especially helped by PCA reducing to an eigenanalysis for which efficient algorithms are readily available. The optimization problem defined by SCoT is more complex. Details of our computational procedures and experience are described in Section 4.

Remark 5. To ensure that successive SCs are different from each other we have imposed the orthogonality constraints $\mathbf{c}_j'\mathbf{c}_k = 0, j < k$, or the constraints $\mathbf{c}_j'\mathbf{S}\mathbf{c}_k = 0, j < k$, which mean that the SCs are pairwise uncorrelated. For PCA the constraints $\mathbf{a}_j'\mathbf{a}_k = 0$ and $\mathbf{a}_j'\mathbf{S}\mathbf{a}_k = 0$ are equivalent, but after rotation at least one of these two properties (orthogonality or uncorrelatedness) is lost (Jolliffe 1995). Similarly, with SCoT we cannot keep both properties. Results for each set of constraints are discussed for the examples which follows.

3. EXAMPLES

3.1 EXAMPLE 1

3.1.1 Simplified Components With Orthogonality

Returning to the example introduced in Section 1, we first use SCoT with the orthogonality constraints, $\mathbf{c}_j'\mathbf{c}_k = 0, j < k$. Table 2 and Figure 1 give loadings of the first four components for PCA, for rotated PCA (RPCA) and for SCoT with $\phi = 1.00, 1.50, 3.0,$ and 6.0. For $\phi = 1.00$ the first SC is very similar to the first PC, though slightly simpler. The second SC is intermediate in simplicity to the second PC and to the second RPC, though "second" has no clear meaning in the case of RPCA. The third and fourth SCs are simpler than either the corresponding PC or RPC.

Table 2. Loadings for the PCA, for RPCA, and for SCoT with varying values of ϕ, for Rencher's meteorological data.

Technique	Component variable	(1)	(2)	(3)	(4)
		Loading matrix			
PCA	x_1	.33	.08	−.09	−.28
(= SCoT	x_2	.35	− .19	−.11	−.23
with	x_3	.39	−.05	−.11	−.14
$\phi = 0$)	x_4	.38	−.05	−.13	−.01
	x_5	.23	−.53	−.01	−.07
	x_6	.36	−.24	−.12	.13
	x_7	−.09	−.02	−.79	.54
	x_8	−.25	−.50	−.08	−.15
	x_9	−.31	−.36	−.21	−.23
	x_{10}	−.02	−.47	.47	.50
	x_{11}	.34	.11	.18	.45
RPCA	x_1	.36	.05	−.12	.23
	x_2	.46	−.10	−.07	.03
	x_3	.43	.05	−.02	.06
	x_4	.39	.10	.08	−.00
	x_5	.39	−.30	−.03	−.31
	x_6	.38	.03	.16	−.21
	x_7	−.01	−.03	.96	.03
	x_8	−.01	−.56	.00	−.17
	x_9	−.05	−.56	.06	.04
	x_{10}	−.13	.02	−.06	−.82
	x_{11}	.09	.50	.10	−.32
SCoT	x_1	−.33	−.06	.02	−.03
($\phi = 1.00$)	x_2	−.35	.15	.03	−.11
	x_3	−.40	.03	.04	−.06
	x_4	−.39	.03	.06	−.04
	x_5	−.23	.60	.01	−.24
	x_6	−.36	.19	.07	−.06
	x_7	.08	.01	.99	.08
	x_8	.24	.51	−.01	−.27
	x_9	.31	.31	.02	−.22
	x_{10}	.02	.45	−.08	.87
	x_{11}	−.33	−.09	.01	.18

Table 2. Continued

Technique	Component variable	(1)	(2)	(3)	(4)
		Loading matrix			
SCoT	x_1	−.33	−.04	.01	.06
($\phi = 1.50$)	x_2	−.36	.04	.04	.14
	x_3	−.40	.00	.04	.10
	x_4	−.39	.00	.05	.10
	x_5	−.22	.17	.04	.21
	x_6	−.36	.07	.07	.14
	x_7	.08	−.04	.99	−.05
	x_8	.24	.13	.02	.92
	x_9	.30	.07	.03	.09
	x_{10}	.02	.97	.02	−.19
	x_{11}	−.33	.01	.01	−.02
SCoT	x_1	−.32	.01	−.02	.07
($\phi = 3.00$)	x_2	−.36	−.02	−.03	.04
	x_3	−.41	−.00	−.03	.07
	x_4	−.40	−.00	−.04	.07
	x_5	−.22	−.08	−.02	−.96
	x_6	−.37	−.03	−.05	.03
	x_7	.08	.02	−1.00	.01
	x_8	.23	−.06	.00	−.14
	x_9	.30	−.03	−.00	−.13
	x_{10}	.02	−.99	−.02	.08
	x_{11}	−.33	−.02	−.02	.10
SCoT	x_1	.09	−.03	.01	−.01
($\phi = 6.00$)	x_2	.10	−.03	−.00	−.01
	x_3	.97	.10	−.01	.02
	x_4	.10	−.04	.00	−.00
	x_5	.06	−.02	−.03	−.00
	x_6	.09	−.04	−.01	.00
	x_7	−.02	.01	.01	1.00
	x_8	−.06	.04	−.03	.01
	x_9	−.07	.05	−.02	.02
	x_{10}	−.01	−.00	−1.00	.01
	x_{12}	.08	−.99	.00	.02

These changes in simplicity are quantified in Table 3, as are the changes in variance. At $\phi = 1.00$ the third SC is almost as simple as it can be, with its varimax value close to 1. SC4, although less simple, has one loading much larger than all others. By contrast, SC2, although simpler than PC2, still has nontrivial loadings on four variables. In terms of simplicity, as measured by the varimax criterion, SC2 is almost as good as RPC2. Its variance is less than that of RPC2, but this is expected because SC1 accounts for a much greater proportion of variance than RPC1. The total variance accounted for by the first two SCs is only slightly less than for the first two PCs. SC2 is a simplified version of PC2, and SC2 has high loadings on a slightly different set of variables compared to RPC2—evaporation (RPC2) is replaced by total wind (SC2) and minimum soil temperature becomes more important (SC2) at the expense of integrated area under the daily humidity curve. These different groupings may provide different insights into the structure of the dataset.

As ϕ increases from 1.00 to 1.50, SC2 in turn becomes dominated by a single variable, and there is corresponding dramatic drop in the variance of SC2. The transition from the form of SC2 seen at $\phi = 1.00$ and that at $\phi = 1.50$ is not smooth. Between $\phi = 1.008$ and $\phi = 1.009$ one local maximum of criterion (2.3) takes over from another as the global maximum, and there is a sudden switch between solutions.

Eventually, by the time ϕ reaches 6.0, all SCs, including the first, represent single variables.

Another way of comparing the results of PCA and SCoT is by plotting the values of the components (scores) for each of the 46 days. This is done in Figures 2–4 for

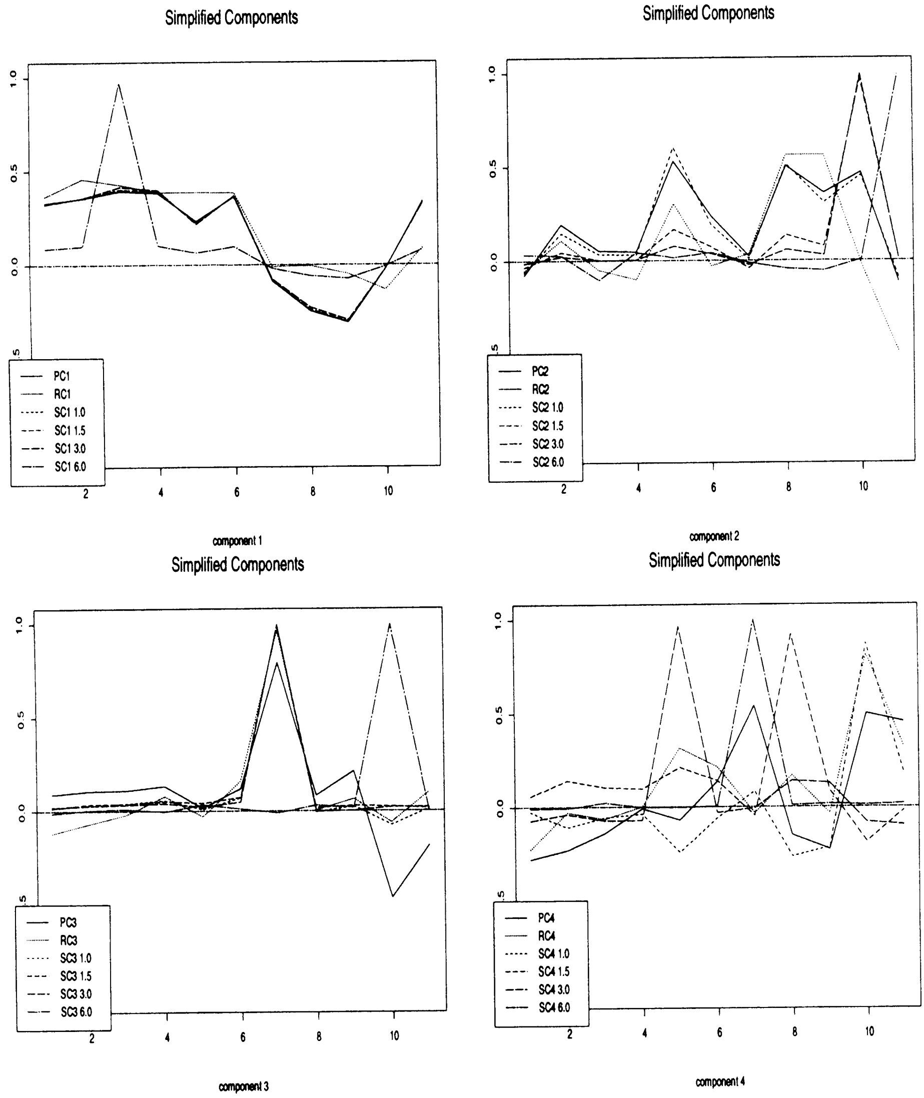

Figure 1. Loadings for the First Four Components of PCA, RPCA, and SCoT with ϕ =1.0, 1.5, 3.0, 6.0 Using Rencher's Meteorological Data Based on the varimax Index of Simplicity. Loadings on vertical axes are plotted against variable labels on horizontal axes.

Table 3. Varimax Index, and Variance for Individual Components in PCA, RPCA, and SCoT for 4 Values of ϕ, for Rencher's Meteorological Data

		Component			
Technique	*Measure*	*(1)*	*(2)*	*(3)*	*(4)*
PCA	varimax index	.03	.03	.40	.12
(= SCoT with	variance	6.02	2.12	1.13	.76
ϕ = 0.00)	Total variance (%)	54.73	74.00	84.27	91.18
RPCA	varimax index	.08	.20	.85	.42
	variance	4.97	2.59	1.02	1.45
	Total variance (%)	45.14	68.73	77.98	91.18
SCoT	varimax index	.03	.18	.96	.55
(ϕ = 1.00)	variance	6.02	2.09	1.01	.72
	Total variance (%)	54.72	73.73	82.92	89.47
SCoT	varimax index	.03	.87	.97	.69
(ϕ = 1.50)	variance	6.02	1.35	.99	.94
	Total variance (%)	54.71	66.99	75.98	84.50
SCoT	varimax index	.04	.97	.98	.85
(ϕ = 3.00)	variance	6.01	1.15	.97	.83
	Total variance (%)	54.63	65.07	73.90	81.43
SCoT	varimax index	.88	.95	.99	1.00
(ϕ = 6.00)	variance	2.22	1.19	1.07	1.00
	Total variance (%)	20.19	31.04	40.76	49.86

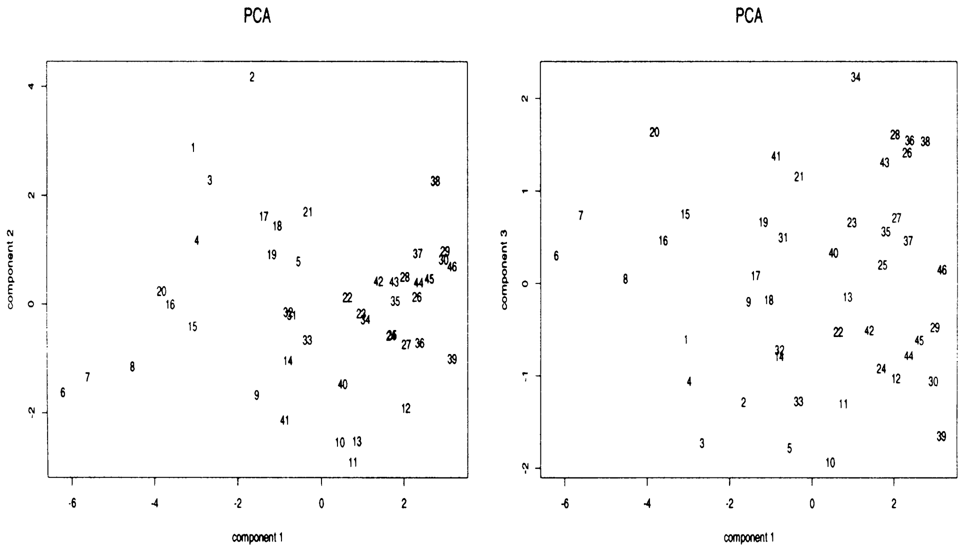

Figure 2. Plots of the First Three PC Scores for Rencher's Meteorological Data.

components 1 to 3. From Figures 2–4, we see that for small values of ϕ, the plots for SCoT are similar, apart from some sign changes, to those of PCA, but as expected there is greater divergence for large values of ϕ. The pattern which is apparent in SC3 for $\phi = 3.0$ is because this component is dominated by variable 7 which takes only a few discrete values.

Rotated components are not plotted. In four-dimensional space, the configuration of the points will be the same as that of PCA, but will be rotated. Because of the arbitrariness of the labeling of the RPCs, it is not meaningful to compare their plots, 2 components at time, with those of PCA and SCoT.

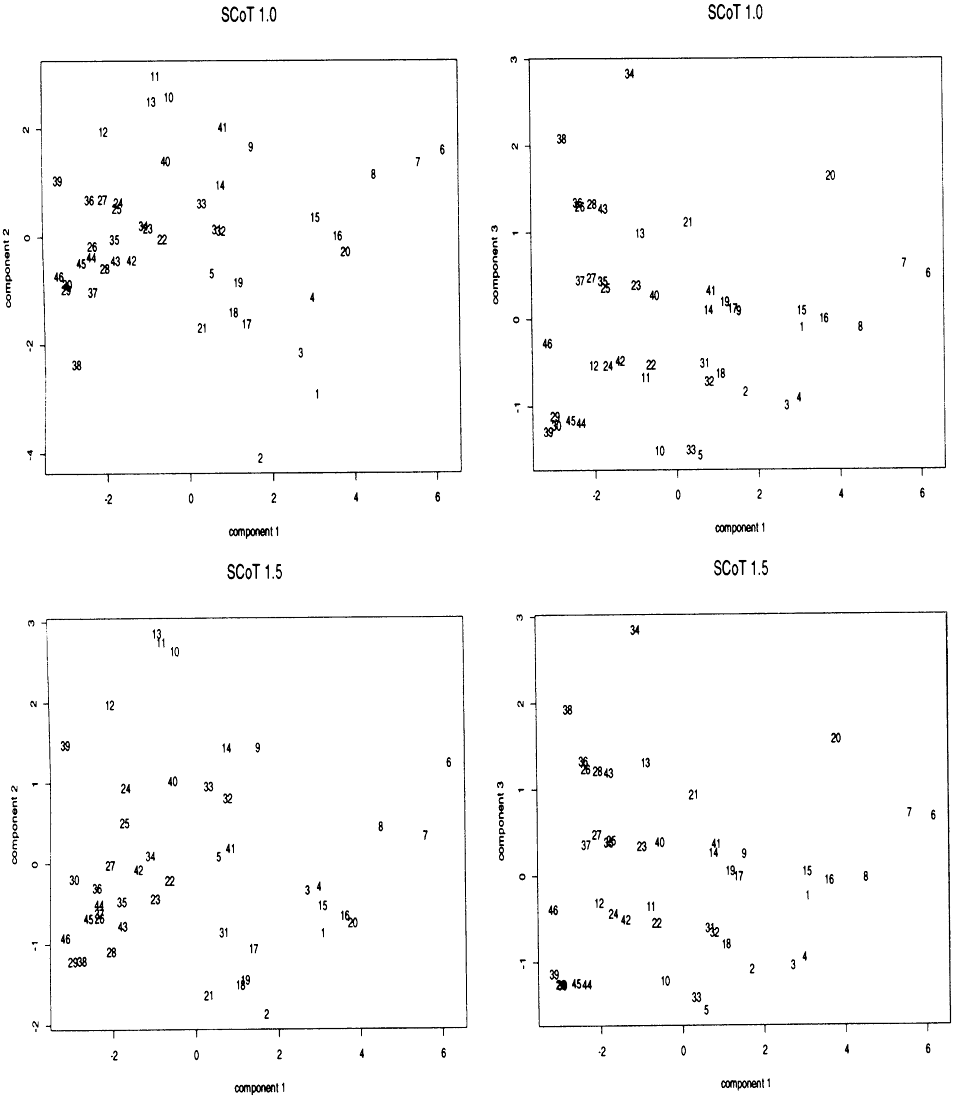

Figure 3. Plots of the First Three SC Scores with $\phi = 1.0, 1.5$ for Rencher's Meteorological Data Based on the varimax Index of Simplicity.

3.1.2 Simplified Components Keeping Uncorrelatedness

It was noted in Section 2.3 that, unlike PCA, we cannot achieve both uncorrelatedness and orthogonality between the simplified components. In this section we report the results for the meteorological data when uncorrelatedness, rather than orthogonality, is imposed. Table 4 gives loadings, simplicity factors and variances for the first four SCs when $\phi = 1.0$. Comparisons with Tables 2–3 show that there is little difference from the orthogonal SCs. The first SC is necessarily the same for both sets of components. As

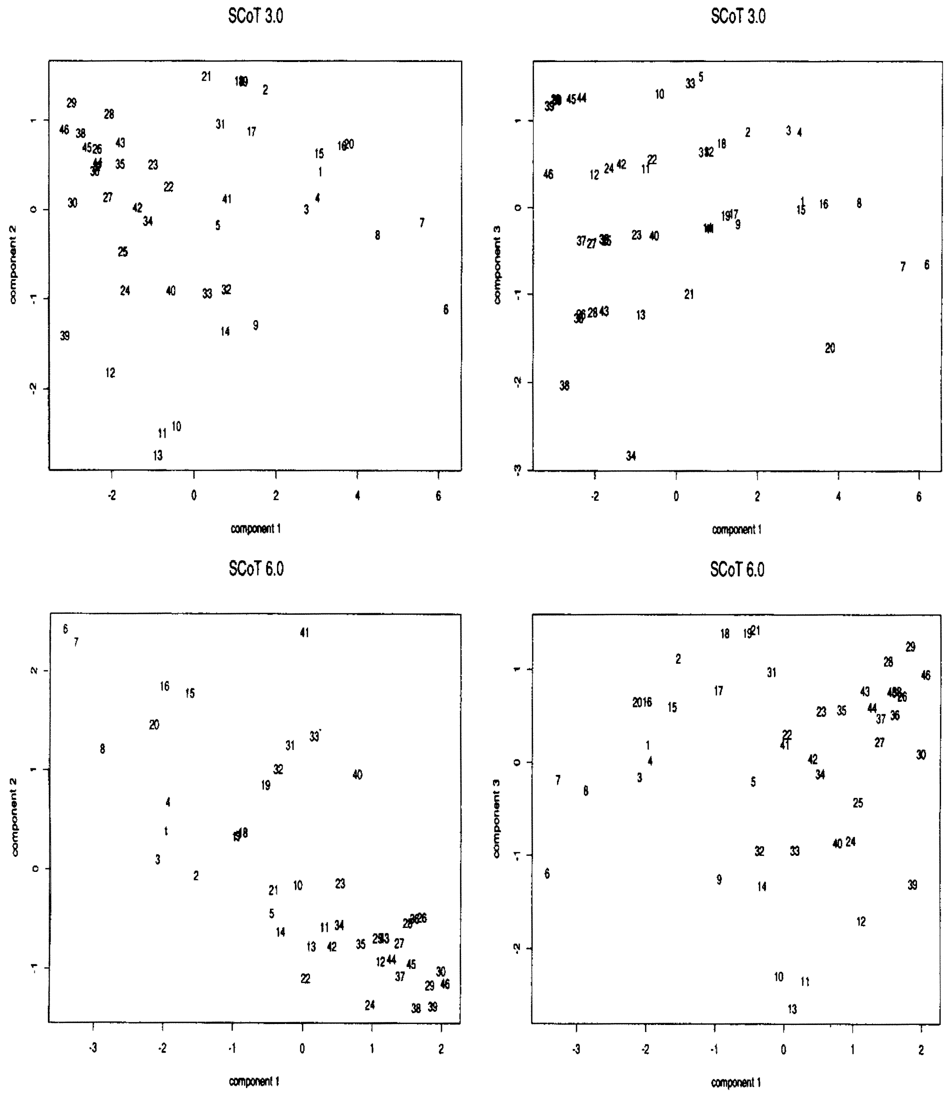

Figure 4. Plots of the First Three SC Scores with ϕ =3.0,6.0 for Rencher's Meteorological Data Based on the varimax Index of Simplicity.

Table 4. Simplified Loadings for Rencher's (1995) Meteorological Data Using Uncorrelated SCoT with $\phi = 1.0$

SC variables	(1)	(2) Loading matrix	(3)	(4)
x_1	.33	.07	−.03	.93
x_2	.35	−.15	−.03	−.08
x_3	.40	−.03	−.04	−.09
x_4	.39	−.03	−.06	−.12
x_5	.23	−.61	−.01	−.02
x_6	.36	−.19	−.07	−.12
x_7	−.08	−.01	−.99	−.00
x_8	−.24	−.51	.01	.12
x_9	−.31	−.31	−.02	.17
x_{10}	−.02	−.45	.08	−.01
x_{11}	.33	.09	−.01	−.19
Simplicity factor	.03	.18	.96	.74
Total Variance (%)	54.72	73.73	82.91	86.64

ϕ increases, the second uncorrelated SC is very similar to the second orthogonal SC until $\phi = 4.0$. As ϕ increases from 4.0, the rate of decrease in variance is slightly higher in the uncorrelated SC than in the orthogonal SC. On the other hand, the second uncorrelated SC achieves greater simplicity than the second orthogonal SC. For SC3 the uncorrelated component has slightly higher variance than the orthogonal component. The fourth uncorrelated component has a higher rate of decrease in variance than that obtained by keeping orthogonality. In contrast to orthogonal SCoT, uncorrelated SCoT will not produce all components equivalent to single variables as ϕ increases (unless the variables themselves are uncorrelated). Tables 3–4 show that SCoT (with either orthogonality and uncorrelatedness) simplifies the components better than RPCA in terms of the varimax criterion. The first four uncorrelated SCs account for 86.64% of the total variance, a roughly 3% reduction in variance compared to orthogonal SCoT, but with a gain in simplicity for SC4. The first three SCs are very similar and, on the whole, the two approaches of the SCoT, keeping orthogonality and maintaining uncorrelatedness, capture similar patterns among the retained sets of the loading matrices. This is not particularly surprising, as orthogonal SCoT produced only slightly correlated components for this example (Table 5). Similarly, the uncorrelated SCoT gives a nearly orthogonal set of vectors.

Table 5. Correlation Matrix for the First Four Orthogonal Simplified Components with $\phi = 1.0$ for Rencher's Meteorological Data

	Correlation Matrix			
Components	(1)	(2)	(3)	(4)
(1)	1.00	.00	−.01	.00
(2)		1.00	.01	.03
(3)			1.00	−.21
(4)				1.00

Table 6. Simplified Loadings for Rencher's Meteorological Data Using SCoT with Variable ϕ = (5.0,1.0,.0001,.0001) Along the First Four Dimensions

SC variables	(1)	(2) Loading matrix	(3)	(4)
$\mathbf{x}_1$	.31	−.07	−.07	.27
$\mathbf{x}_2$	.35	.14	−.10	.22
$\mathbf{x}_3$	.47	.03	−.11	.14
$\mathbf{x}_4$	.41	.03	−.12	.01
$\mathbf{x}_5$	.20	.58	−.01	.06
$\mathbf{x}_6$	.36	.18	−.11	−.14
$\mathbf{x}_7$	−.07	.01	−.79	−.55
$\mathbf{x}_8$	−.21	.53	−.11	.16
$\mathbf{x}_9$	−.28	.32	−.24	.23
$\mathbf{x}_{10}$	−.02	.45	.45	−.49
$\mathbf{x}_{11}$	.31	−.10	.21	−.46
Simplicity factor	.05	.17	.39	.13
Total Variance (%)	54.13	73.22	83.58	90.50

3.1.3 Simplified Components with Variable ϕ

The behavior in this, and other, examples indicates that if we take ϕ large enough for SC1 to become noticeably different from PC1, all the subsequent SCs are often dominated by single variables. Allowing the possibility of using different values of ϕ for different SCs may avoid this and will increase flexibility. A smaller value of ϕ will be used in deriving SC2 than for SC1, a still smaller value for SC3 and so on.

Table 6 gives results for the meteorological data set with $\phi = (5.0, 1.0, .0001, .0001)$. As before, the transition between different patterns in the SCs with the increase in ϕ is not at all smooth, and to obtain a compromise for all SCs between closeness to PCA and domination by a single variable, the values of ϕ for different SCs are very different. Compared to PCA, the third and fourth components now gain only a very small amount of simplicity at the cost of losing a small amount of variance (about 1%) along the retained set of first four components. For this reason the retained components are nearly uncorrelated and remain close to PCA (Tables 2, 6, and 7).

Table 7. Correlation Matrix for the First Four Orthogonal Simplified Components for Rencher's Meteorological Data Using SCoT with Variable ϕ = (5.0,1.0,.0001,.0001) Along the First Four Dimensions

	Correlation matrix			
Components	(1)	(2)	(3)	(4)
(1)	1.00	−.03	.08	−.03
(2)		1.00	−.02	−.00
(3)			1.00	.00
(4)				1.00

Table 8. Definitions of Variables in Jarvik's Smoking Questionnaire

Variable	*Definition*
x_1	Concentration
x_2	Annoyance
x_3	Smoking 1 (first wording)
x_4	Sleepiness
x_5	Smoking 2 (second wording)
x_6	Tenseness
x_7	Smoking 3 (third wording)
x_8	Alertness
x_9	Irritability
x_{10}	Tiredness
x_{11}	Contentedness
x_{12}	Smoking 4 (fourth wording)

3.2 EXAMPLE 2

The data analyzed in this section are from Jarvik's smoking questionnaire (M. E. Jarvik, unpublished). The data involve the answers by 110 individuals to 12 questions, listed in Table 8, each coded from 1 to 5 such that a high score represents a desire to smoke. The four questions labeled "smoking" are direct questions about the subject's desire to have a cigarette. The remainder of the questions relate to the psychological and physical state of the subjects.

Figure 5 gives loadings for PCA, for RPCA, and for SCoT with a range of values of ϕ, and Table 9 contains information on the variances and the values of the varimax criterion for each component in Figure 5. The behavior seen in Figure 5 and Table 9 is similar to that in Figure 1 and Table 3. The first SC remains similar to PC1 until ϕ is quite large, and as ϕ increases SC3, then SC2 and finally SC1 becomes dominated by a single variable.

Examining, for example, SCoT when $\phi = 2.0$ we find that SC1 is still close to PC1 and SC3 is dominated by a single variable, Smoking 1 (First wording). SC2 has its highest loadings on the same four variables as RPCA but is simpler in the sense that one variable (contentedness) plays a much more dominant role compared to other three variables than in RPCA. Compared to PCA, the first two SCs lose 9%, while RPCA loses 18%. With three components, where RPCA necessarily retains the same total variability as PCA, SCA loses 13% but with a considerable gain in simplicity.

4. SIMULATION STUDIES

One question of interest is whether SCoT is better at detecting underlying simple structure in a dataset when it exists than is PCA or RPCA. To investigate these questions we simulated data from a variety of known structures. Only a small part of the results is given here; further details can be found in Uddin (1999).

4.1 Structure of the Simulated Data

Given a vector $\boldsymbol{l}$ of positive real numbers and an orthogonal matrix $\mathbf{A}$, we can attempt to find a covariance matrix or correlation matrix whose eigenvalues are the elements of $\boldsymbol{l}$, and whose eigenvectors are the columns of $\mathbf{A}$. Some restrictions need to be imposed on $\boldsymbol{l}$ and $\mathbf{A}$, especially in the case of correlation matrices, but it is possible to find such matrices for a wide range of eigenvector structures. Having obtained a covariance or correlation matrix it is straightforward to generate samples of data from multivariate

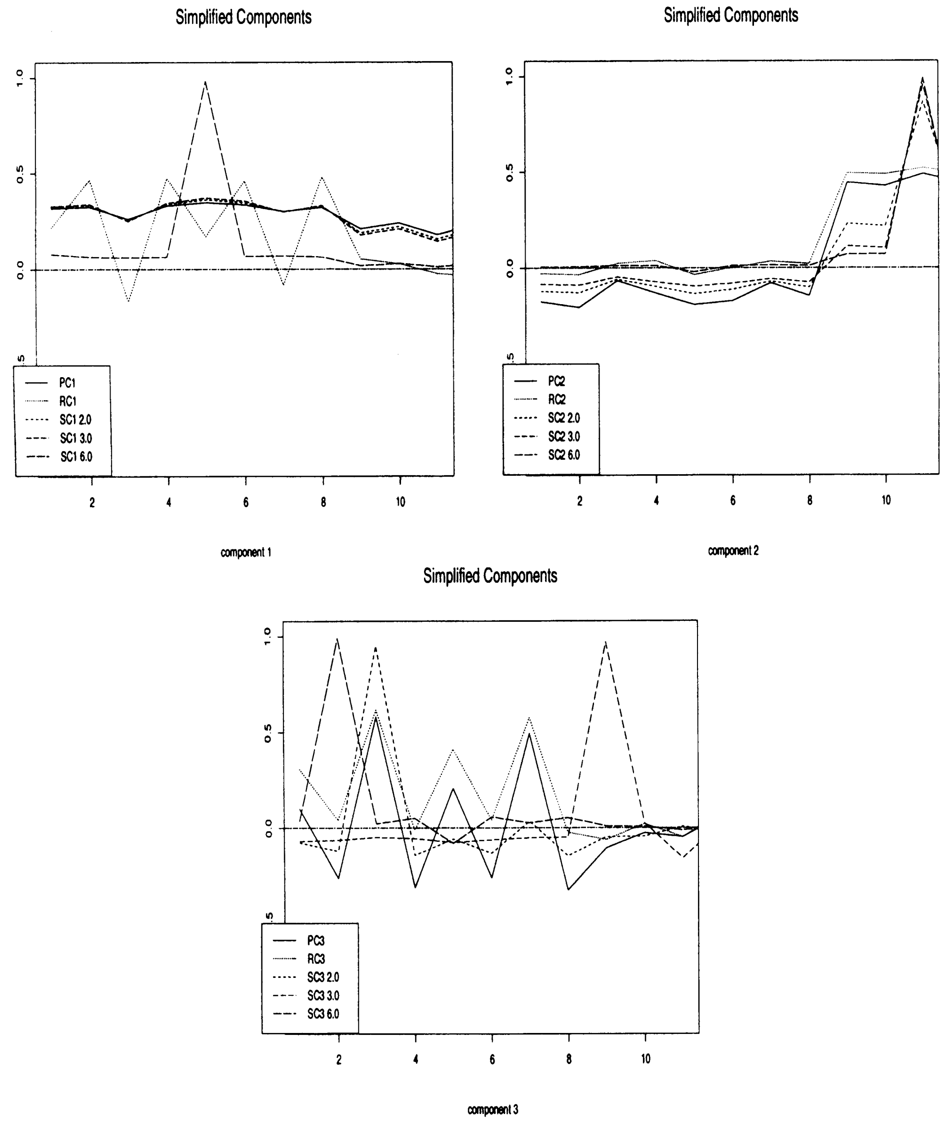

Figure 5. Loadings for the First Three Components of PCA, RPCA, and SCoT with ϕ =2.0, 3.0, 6.0 using Jarvik's Smoking Data Based on the varimax Index of Simplicity. Loadings on vertical axes are plotted against variable labels on horizontal axes.

Table 9. Varimax Index, and Variance for Individual Components in PCA, RPCA, and SCoT for 3 Values of ϕ, for Jarvik's Smoking Data

Technique	*Measure*	*Components* (1)	(2)	(3)
PCA	varimax index	.01	.09	.14
(= SCoT with	variance	5.43	3.00	1.36
$\phi = 0$)	Total variance (%)	45.21	70.19	81.52
RPCA	varimax index	.13	.18	.23
	variance	3.44	3.44	2.89
	Total variance (%)	28.69	57.39	81.52
SCoT	varimax index	.02	.54	.81
($\phi = 2.00$)	variance	5.42	2.23	.83
	Total variance (%)	45.16	63.77	70.72
SCoT	varimax index	.02	.82	.88
($\phi = 3.00$)	variance	5.41	1.59	.78
	Total variance (%)	45.08	58.34	64.86
SCoT	varimax index	.93	.96	.96
($\phi = 6.00$)	variance	1.74	1.39	1.23
	Total variance (%)	14.47	26.10	36.36

normal distributions with the given covariance or correlation matrix. We have done this for a variety of eigenvector structures (principal component loadings), and computed the PCs, RPCs, and SCs from the resulting sample correlation matrices. Three main types of structures have been investigated, which we call block structure, intermediate structure and uniform structure. Tables 10–12 give one example of each type of structure.

In "uniform structures," some or all of the underlying PCs have roughly equal-sized loadings (in absolute value) for *all* variables; "block structures" have *subsets* (blocks) of important variables in each PC; "intermediate" structures lie between these two patterns.

4.2 RESULTS

We give one (out of many) set of results for the each of the three structures in Tables 10–12. Here a sample of 75 observations is generated from each structure and the results

Table 10. Specified Eigenvectors of a 6-dimensional Block Structure

Variable	*Eigenvectors* (1)	(2)	(3)	(4)	(5)	(6)
x_1	.10	−.54	.76	−.12	.33	−.02
x_2	.08	−.56	−.60	.23	.51	−.01
x_3	.08	−.60	−.12	−.12	−.77	.02
x_4	.59	.08	−.07	−.31	.07	.73
x_5	.58	.10	−.11	−.42	.05	−.68
x_6	.53	.07	.18	.80	−.16	−.07
Variance	1.84	1.64	.75	.66	.61	.51

Table 11. Specified Eigenvectors of a 6-dimensional Intermediate Structure

	Eigenvectors					
Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
x_1	.22	−.51	.60	.30	−.33	.36
x_2	.25	−.52	−.36	−.64	−.34	−.06
x_3	.23	−.55	−.25	.38	.61	−.27
x_4	.55	.25	−.25	−.05	.26	.71
x_5	.52	.25	−.26	.45	−.51	−.37
x_5	.51	.20	.56	−.38	.28	−.40
Variance	1.79	1.67	.80	.62	.61	.51

of PCA, RPCA, and SCoT are recorded. Results similar to those now presented were obtained for other replications from the same structures, for other structures within the same broad categories, and for 10 variables rather than 6.

It might be expected that if the underlying structure is simple, then sampling variation is more likely to take sample PCs away from simplicity than to enhance this simplicity. It is of interest to investigate whether the techniques of RPCA and SCoT which increase simplicity compared to the sample PCs will do so in the direction of the true underlying structure. The closeness of a vector of loadings from any of these techniques to the underlying true vector is measured by the angle between the two vectors of interests. These angles are given in Tables 13–15.

The results illustrate that, for each structure, RPCA is perhaps surprisingly, and certainly disappointingly, bad at recovering the underlying structure. SCoT, on the other hand, is capable of improvement over PCA. For example, in Table 14 for $\phi = .18$ it improves over PCA in terms of angles in Table 14, thus giving a notably simpler structure. None of the methods manages to reproduce the underlying structure for component 4 in Tables 13 and 14. For both block and uniform structure SCoT remains close to PCA (Tables 13 and 15). It is perhaps surprising that SCoT does less well for both block and uniform structure than for intermediate structure. This behavior was repeated in other replications and we have no explanation for it.

Similar simulation studies have been done for covariance matrices and here the pattern of results is rather different, with SCoT showing greatest improvement over PCA for block structure and least for uniform structure. RPCA is poor for all structures. In

Table 12. Specified Eigenvectors of a 6-dimensional Uniform Structure

	Eigenvectors					
Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
x_1	−.45	.34	−.09	.74	−.33	.12
x_2	−.44	.37	−.21	−.63	−.44	−.17
x_3	−.41	.42	.38	−.11	.70	.10
x_4	.43	.46	.04	−.14	−.17	.74
x_5	.30	.43	−.70	.11	.36	−.30
x_6	.38	.42	.56	.10	−.23	−.54
Variance	1.84	1.71	.80	.65	.52	.48

Table 13. Angles Between the Underlying Vectors and the Sample Vectors of PCA, RPCA, and SCoT with Various Values of ϕ, for a Specified Block Structure of Correlation Eigenvectors

		Vectors			
Technique	ϕ	*(1)*	*(2)*	*(3)*	*(4)*
PCA	0	12.9	12.0	15.4	79.5
RPCA		37.7	45.2	45.3	83.1
SCoT	.11	12.5	11.8	15.7	70.6
	.18	12.4	11.9	17.0	69.3
	.25	12.2	12.2	19.4	68.8
	.33	12.1	13.0	23.8	68.9
	.43	12.0	15.0	31.6	69.6

both correlation and covariance simulations much smaller values of ϕ, compared to those used in the examples of Section 3, are needed to show improvement for SCoT over PCA. We return to this point in Section 6.

Finally in this section, we note that comparisons of underlying and sample structures are complicated by switching of the order of the first two components in some replications. This switching may occur because of a change of signs in some blocks within the population and sample correlation or covariance matrices, and is sometimes due to the closeness of the first two population eigenvalues; there are occasions when both phenomena are present.

5. COMPUTATION

5.1 OPTIMIZATION

The simplified component technique relies on numerical optimization to estimate parameters and it suffers from the problem of many local optima. Traditional methods of function optimization can have considerable difficulty in converging to the global optimum when multiple optima are present, as is the case here. In contrast to traditional

Table 14. Angles Between the Underlying Vectors and the Sample Vectors of PCA, RPCA, and SCoT with Various Values of ϕ, for a Specified Intermediate Structure of Correlation Eigenvectors

		Vectors			
Technique	ϕ	*(1)*	*(2)*	*(3)*	*(4)*
PCA	0	13.7	15.1	23.6	78.3
RPCA		53.3	42.9	66.4	77.7
SCoT	.05	10.7	12.5	22.7	77.8
	.11	6.3	9.2	21.4	76.9
	.18	6.4	9.2	19.8	75.2
	.25	10.0	12.0	17.5	71.3
	.33	13.2	14.8	15.9	66.1
	.43	16.2	17.6	21.7	64.1

Table 15. Angles Between the Underlying Vectors and the Sample Vectors of PCA, RPCA, and SCoT with Various Values of ϕ, for a Specified Uniform Structure of Correlation Eigenvectors

		Vectors			
Technique	ϕ	*(1)*	*(2)*	*(3)*	*(4)*
PCA	0	6.0	6.4	23.2	31.8
RPCA		52.0	54.1	39.4	43.5
SCoT	.11	6.0	6.7	21.5	24.9
	.18	6.1	6.9	21.3	22.5
	.25	6.2	7.2	21.7	21.4
	.33	6.2	7.5	22.5	22.4
	.43	6.4	8.0	23.9	26.1

optimization methods, simulated annealing is less likely to fail in finding the global optima of a multidimensional function (Lundy and Mees 1986; Mitra, Romeo, and Sangiovanni-Vincentelli 1986; Goffe, Ferrier, and Rogers 1994). However, the main drawback of the simulated annealing method, which has limited its use, is that convergence can be very slow. These algorithms can be speeded up, but at the expense of not always finding a *global* optimum.

For the case of multiple optima, a common approach is to run traditional routines from a number of starting points and to choose the best solution from those obtained. The main difficulty with this approach arises with the choice of starting points as traditional methods are sensitive to the starting points. Since we have no a priori knowledge of the p-dimensional surface, this choice is arbitrary and has a considerable effect upon the reliability and efficiency of the algorithm. However, reliability can be achieved to some extent by iterating the underlying function for a sufficient number of times for different well-spaced starting points.

In this study, we developed a new technique which produces a set of simplified components. In this technique, we used the built-in quasi-Newton function of S-Plus, interfaced with the modified FORTRAN functions based on **dmnfb**, **dmngb**, and **dmnhb** (Gay 1983, 1984). The modified FORTRAN functions were used for both constrained and unconstrained maximization in the proposed S-Plus code of SCoT, and depend on whether we used variable or fixed ϕ. To achieve an appropriate solution, several different random starting points were used along each dimension and the quasi-Newton method was applied. The maximum value of the objective function defined in (2.3) and the corresponding vectors thus obtained were selected.

To ensure the criterion values for SC1, SC2, SC3, ... were in decreasing order, and to ensure that all the maxima were global maxima, the problem was solved by replicating the optimization 11 times, with different randomly chosen unit vectors and for different starting values. Adopting this strategy, the convergence is quicker than the simulated annealing method.

5.2 ORTHOGONALITY

Another important aspect of the SCoT is the implementation of orthogonality or uncorrelatedness. In the sequential approach of PCA, orthogonality between the successive vectors $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$ can be maintained by adjusting the matrix of residual correlations or variances and covariances.

In SCoT, the Gram-Schmidt process of orthogonalization was used during the maximization so that the successive vectors $(\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m)$ were orthogonal. We find the first vector $\mathbf{c}_1$ after maximizing the objective function defined in (2.3). The successive vectors $\mathbf{c}_k$ were obtained by spinning arbitrarily chosen vectors $\boldsymbol{\alpha}_k$ under maximization by using the transformation

$$\begin{aligned}
\mathbf{c}_2 &= \boldsymbol{\alpha}_2 - \left(\frac{\boldsymbol{\alpha}_2'\mathbf{c}_1}{\mathbf{c}_1'\mathbf{c}_1}\right)\mathbf{c}_1 \\
\mathbf{c}_3 &= \boldsymbol{\alpha}_3 - \left(\frac{\boldsymbol{\alpha}_3'\mathbf{c}_2}{\mathbf{c}_2'\mathbf{c}_2}\right)\mathbf{c}_2 - \left(\frac{\boldsymbol{\alpha}_3'\mathbf{c}_1}{\mathbf{c}_1'\mathbf{c}_1}\right)\mathbf{c}_1 \\
\vdots &= \vdots \\
\mathbf{c}_k &= \boldsymbol{\alpha}_k - \left(\frac{\boldsymbol{\alpha}_k'\mathbf{c}_{k-1}}{\mathbf{c}_{k-1}'\mathbf{c}_{k-1}}\right)\mathbf{c}_{k-1} - \cdots - \left(\frac{\boldsymbol{\alpha}_k'\mathbf{c}_1}{\mathbf{c}_1'\mathbf{c}_1}\right)\mathbf{c}_1,
\end{aligned} \tag{5.1}$$

where the first vector $\mathbf{c}_1$ was considered as fixed, in finding the successive set of vectors during the Gram-Schmidt orthogonalization. For each $\mathbf{c}_k$, $k \geq 2$, each iteration within the quasi-Newton method is followed by a Gram-Schmidt orthogonalization step.

5.3 UNCORRELATEDNESS

The procedure described in Section 5.2 can be adapted to ensure uncorrelatedness, rather than orthogonality, between SCs. In this case we require the constraints $\mathbf{c}_h'\mathbf{S}\mathbf{c}_k = 0$, $h \neq k$, instead of $\mathbf{c}_h'\mathbf{c}_k = 0$ for orthogonality.

To maintain uncorrelatedness between the SCs, the Gram-Schmidt process of orthogonalization can once again be used. The condition $\mathbf{c}_h'\mathbf{S}\mathbf{c}_k = 0$, is equivalent to $\mathbf{y}_h'\mathbf{y}_k = 0$, where $\mathbf{y}_k, k = 1, 2, \ldots$ is the kth SC, defined as $\mathbf{y}_k = \mathbf{X}\mathbf{c}_k$, where $\mathbf{X}$ is the $(n \times p)$ column-centered matrix of the n observations on p variables. Hence $\mathbf{c}_h'\mathbf{S}\mathbf{c}_k = 0$ is equivalent to "orthogonality" between $\mathbf{y}_h$ and $\mathbf{y}_k$.

We first find the vector $\mathbf{c}_1$ by maximizing the objective function defined in (2.3), and this yields the first SC, $\mathbf{y}_1 = \mathbf{X}\mathbf{c}_1$.

The successive SCs are obtained by spinning the arbitrarily chosen components $\boldsymbol{\zeta}_k = \mathbf{X}\boldsymbol{\alpha}_k$ (based on the arbitrarily chosen vectors $\boldsymbol{\alpha}_k$) under maximization by using the transformation

$$
\begin{aligned}
\mathbf{y}_2 &= \zeta_2 - \left(\frac{\zeta_2'\mathbf{y}_1}{\mathbf{y}_1'\mathbf{y}_1}\right)\mathbf{y}_1 \\
\mathbf{y}_3 &= \zeta_3 - \left(\frac{\zeta_3'\mathbf{y}_2}{\mathbf{y}_2'\mathbf{y}_2}\right)\mathbf{y}_2 - \left(\frac{\zeta_3'\mathbf{y}_1}{\mathbf{y}_1'\mathbf{y}_1}\right)\mathbf{y}_1 \\
\vdots &= \vdots \\
\mathbf{y}_k &= \zeta_k - \left(\frac{\zeta_k'\mathbf{y}_{k-1}}{\mathbf{y}_{k-1}'\mathbf{y}_{k-1}}\right)\mathbf{y}_{k-1} - \cdots - \left(\frac{\zeta_k'\mathbf{y}_1}{\mathbf{y}_1'\mathbf{y}_1}\right)\mathbf{y}_1,
\end{aligned}
\tag{5.2}
$$

where the first SC, $\mathbf{y}_1$, is kept fixed. The vectors corresponding to the uncorrelated components are extracted by the back transformation $\mathbf{c}_k = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_k$. The selected vectors $\mathbf{c}_k$'s have then been normalized so that they are of unit length. For each $\mathbf{y}_k$, $k \geq 2$, each iteration within the quasi-Newton method is followed by a Gram-Schmidt orthogonalization step which brings uncorrelatedness between the successive SCs.

6. DISCUSSION AND CONCLUSION

The results of Section 3, and other examples we have examined, show a common pattern of behavior of SCoT. The simplified components are necessarily close to principal components for small ϕ, and as ϕ increases so individual components become dominated by single variables. Lower-variance components tend to take this simpler form for smaller values of ϕ than higher-variance components. Transition between different solutions as ϕ increases can be sudden, corresponding to switching between different local maxima in the optimization problem. This is noted in the discussion of Example 1, and similar behavior occurs for SC1 in Example 2 between $\phi = 5.40$ and $\phi = 5.41$. The transition to a dominant-single-variable SC2 appears much smoother in Example 2, but even here we switch minima as SC3 goes from being dominated by x_3 to dominance by x_9, as the latter's contribution to SC2 become smaller.

Despite a tendency for SCs to remain close to PCs or become dominated by a single variable, there are typically some values of ϕ for which some SCs have loadings which are different from the PCs and the RPCs, without becoming single-variable components, and without sacrificing a great deal of variance. Such simplified components can give new insights into the underlying structure of the data.

Our examples used correlation matrices, but there is no reason why SCoT should not be used on covariance matrices. The fact that covariance matrix analysis only makes sense when all variables are measured in the same units with a relatively small spread in their variance, means that covariance matrix analyses are much less common than correlation analyses. We have, however, examined SCoT in this context. The results are generally in line with expectation. When variances are roughly equal, there is little difference from the correlation matrix results. As the variances become increasingly different, the early PCs become dominated by single, high variance, variables, so that SCoT does little to change PCA.

A question which arises in a covariance matrix analysis is whether, and how, to "standardize" ϕ so that similar values are relevant to analyses with very different diagonal elements in the covariance matrix. An intuitive idea is to multiply the variance term, the first part of the objective function defined in (2.3), by a factor $p/\text{trace}(\mathbf{S})$, so that this part has the same range of values as for a correlation matrix of the same dimension, p. Other modifications to (2.3) are also possible for both correlation and covariance matrices. We could simply divide the first term by $\text{trace}(\mathbf{S})$, which equals p for correlation matrices. Alternatively we could express (2.3) as a weighted average $\psi V(\mathbf{c}_k) + (1 - \psi)S(\mathbf{c}_k)$, of $V(\mathbf{c}_k)$ and $S(\mathbf{c}_k)$. Another variation of the technique would be to maximize $V(\mathbf{c}_k)$ subject to $S(\mathbf{c}_k) = \tau$, for some fixed τ, a form of PCA with constraints. This will give us SC1, but later components will be SCs with a particular set of variable ϕ's, depending on τ.

The question of the choice of ϕ is unresolved. The values in the simulation studies which gave improvements for SCoT in finding the underlying structures were much smaller than those used in the real examples. Such values simplify later components (SC3, SC4 in the meteorological data), but SC1, SC2 are very similar to PC1, PC2 in our examples.

Any simulation study can necessarily only cover a very small proportion of potential structures and it may be that there are other structures for which ϕ should be larger. Further work is needed, but our recommendation at present is that results for at least two or three values of ϕ should be examined if we wish to find components with different patterns from PCA, but which have not become dominated by single variables. It seems likely that, even with further work, it will rarely be the case that a clear-cut recommendation for a single value of ϕ can be given. The fact that it may be desirable to use quite different ϕ's for different SCs, as in Section 3.1.3, reinforces this view.

In SCoT, as in most optimization problems, the global optimum is required and nonglobal local optima are a nuisance. In our implementation, we try to ensure that all the maxima are global maxima, by replicating the optimization 11 times, with different randomly chosen unit vectors and for different starting values. Adopting this strategy with the quasi-Newton method, the convergence is quicker than the simulated annealing method. However, there is probably scope for improvement in speed and in avoiding local optima.

Finally, we note that as well as the varimax criterion we have implemented SCoT with other orthogonal rotation criteria. Similar results were found throughout our analyses, so we have reported only the varimax results.

ACKNOWLEDGMENTS

The authors are grateful to two anonymous referees, an associate editor, and the editor, Andreas Buja, for valuable criticisms and suggestions, which led to considerable improvements in the article.

[Received December 1998. Revised November 1999.]

REFERENCES

Cadima, J., and Jolliffe, I. T. (1995), "Loadings and Correlations in the Interpretation of Principal Components," *Journal of Applied Statistics*, 22, 203–214.

Gay, D. M. (1983), "Algorithm 611. Subroutines for Unconstrained Optimization Using a Model/Trust-Region Approach," *ACM Transactions on Mathematical Software*, 9, 503–524.

——— (1984), "A Trust Region Approach to Linearly Constrained Optimization," in *Numerical Analysis. Proceedings*, ed. F. A. Lootsma, New York: Springer, pp. 171–189.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994), "Global Optimizations of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalised Linear Models: A Roughness Penalty Approach*, London: Chapman and Hall.

Hausman, R. (1982), "Constrained Multivariate Analysis," in *Optimisation in Statistics*, eds. S. H. Zanckis and J. S. Rustagi, Amsterdam: North Holland, pp. 137–151.

Huber, P. J. (1985), "Projection Pursuit" (with discussion), *The Annals of Statistics*, 13, 435–525.

Jolliffe, I. T. (1987), "Rotation of Principal Components: Some Comments," *Journal of Climatology*, 7, 507–510.

——— (1989), "Rotation of Ill-Defined Principal Components," *Applied Statistics*, 38, 139–147.

——— (1995), "Rotation of Principal Components: Choice of Normalization Constraints," *Journal of Applied Statistics*, 22, 29–35.

Jolliffe, I. T., and Uddin, M. (in press), "A Modified Principal Component Technique Based on the Lasso," submitted to *Journal of Computational and Graphical Statistics*.

Jones, M. C., and Sibson, R. (1987), "What is Projection Pursuit?" (with discussion), *Journal of the Royal Statistical Society*, Series A, 150, 1–36.

Krzanowski, W. J., and Marriott, F. H. C. (1995), *Multivariate Analysis*, Part II, London: Arnold.

Lundy, M., and Mees, A. (1986), "Convergence of an Annealing Algorithm," *Mathematical Programming*, 34, 111–124.

Mitra, D., Romeo, F., and Sangiovanni-Vincentelli, A. L. (1986), "Convergence and Finite-Time Behaviour of Simulated Annealing," *Advances in Applied Probability*, 18, 747–771.

Morton, S. C. (1989), "Interpretable Projection Pursuit," Technical Report 106, Department of Statistics, Stanford University, Stanford, California.

Rencher, A. C. (1995), *Methods of Multivariate Analysis*, New York: Wiley.

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society*, Series B, 53, 233–243.

Richman, M. B. (1986), "Rotation of Principal Components," *Journal of Climatology*, 6, 293–335.

——— (1987), "Rotation of Principal Components: A Reply," *Journal of Climatology*, 7, 511–520.

Titterington, D. M. (1985) "Common Structure of Smoothing Techniques in Statistics," *International Statistical Review*, 53, 141–170.

Uddin, M. (1999), "Interpretation of Results from Simplified Principal Components," unpublished Ph.D. thesis, University of Aberdeen, Aberdeen, Scotland.

Vines, S. K. (2000), "Simple Principal Components," *Applied Statistics*, 49, 441–451.